



## King's Research Portal

### *Document Version*

Early version, also known as pre-print

[Link to publication record in King's Research Portal](#)

### *Citation for published version (APA):*

Sena, A., Michael, B., & Howard, M. J. W. (Accepted/In press). *Improving Task-Parameterized Movement Learning with Frame-Weighted Trajectory Generation*. Paper presented at IEEE/RSJ International Conference on Intelligent Robots and Systems, Macau, China. <https://arxiv.org/abs/1903.01240>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Improving Task-Parameterised Movement Learning Generalisation with Frame-Weighted Trajectory Generation

Aran Sena, Brendan Michael and Matthew Howard

**Abstract**—Learning from Demonstration depends on a robot learner generalising its learned model to unseen conditions, as it is not feasible for a person to provide a demonstration set that accounts for all possible variations in non-trivial tasks. While there are many learning methods that can handle *interpolation* of observed data effectively, *extrapolation* from observed data offers a much greater challenge. To address this problem of generalisation, this paper proposes a modified Task-Parameterised Gaussian Mixture Regression method that considers the *relevance of task parameters* during trajectory generation, as determined by variance in the data. The benefits of the proposed method are first explored using a simulated reaching task data set. Here it is shown that the proposed method offers far-reaching, low-error extrapolation abilities that are different in nature to existing learning methods. Data collected from novice users for a real-world manipulation task is then considered, where it is shown that the proposed method is able to effectively reduce grasping performance errors by  $\sim 30\%$  and extrapolate to unseen grasp targets under real-world conditions. These results indicate the proposed method serves to benefit novice users by placing less reliance on the user to provide high quality demonstration data sets.

## I. INTRODUCTION

This paper considers robot Learning from Demonstration (LfD), in which examples of how to perform a task are collected from a human teacher such that the robot can use them to learn a model to perform the demonstrated task. In particular, it considers a learned model’s ability to perform under situations that were not demonstrated, *i.e.*, the learners ability to *generalise*, and presents a new method that significantly improves task performance in unseen conditions.

A strength often mentioned when introducing LfD is that it enables *novice users*, people who do not have the relevant knowledge to effectively program a robot, to deploy robots in labour intensive tasks by reducing the need for technical expertise [1], [2]. A corresponding weakness is then the inability for any person interacting with the robot to provide demonstrations for all conceivable variations of a non-trivial task. Furthermore, the person teaching is fallible and prone to poor teaching behaviours such as not being able to gauge the appropriate number of demonstrations required for a robot to learn a task and struggling to identify gaps in the learners knowledge [1], [3]. To overcome the limitations of the teacher and adapt to new situations, the robot learner must

be able to effectively generalise from the demonstrations provided.

Generalisation can take two forms, namely (i) *interpolation*, and (ii) *extrapolation*. In the former, the learner must perform the task under conditions that are *within* some range of conditions they have previously observed. In the latter, they must perform the task under conditions that are out-of-range of their observed experience. Many learning methods will perform well under interpolation conditions, but then degrade in performance under extrapolation [4]. Improving a robot learner’s ability to extrapolate would help them to effectively learn tasks from limited demonstrations, and reduce teaching effort for their human users.

One way to improve the extrapolation ability of a learner is to consider the *local structure* present in a task. For example, in learning to pick up a coffee mug, it would be beneficial for the robot to learn that the approach direction of its gripper is important for successfully grasping the mug handle.

This approach of exploiting the local structure is used in a class of methods known as Task Parametrised (TP) learning, as presented in [4].

In TP learning, specific task relevant parameters are defined, such as object positions and orientations in an environment, and these are used to construct frames of reference. Data collected from the robot-point-of-view can then be “observed” from different points of view through these alternative frames of reference. Considering the coffee mug example, from the robot’s perspective the teacher demonstrates how to reach for cups located in different locations, while from the cup’s perspective, the teacher is demonstrating how to approach the cup from many start locations. By learning task representations in these local frames of reference, the learner’s extrapolation abilities can be improved a great deal.

This paper presents a modified Task-Parameterised Gaussian Mixture Regression (TP-GMR) method that considers the *relevance of particular task parameters* before combining them into the global model. By doing so, local structure can be more effectively preserved, resulting in improved task performance. The benefits of the proposed method are shown through two experiments, highlighting the difficulty existing methods have in maintaining the local structure of a demonstrated task under extrapolation conditions. Significant improvement using the proposed method is shown in a test data set for a reaching task, versus the original TP-GMR method described in [4] and a modified TP method described in [5], specifically designed to improve extrapolation of learned skills. Significant improvement is then also shown

Robot Learning Lab, Department of Informatics, King’s College London.  
aran.sena@kcl.ac.uk, aransena@gmail.com

This work was funded by the Agriculture and Horticulture Development Board (AHDB) GROWBOT project, HNS/PO 194, and the Engineering and Physical Sciences Research Council (EPSRC) SoftSkills project, EP/P010202/1.

for a real-world manipulation task, with a  $\sim 30\%$  reduction in task error in a data set of 108 teaching interactions with novice users.

## II. RELATED WORK

Generalisation of demonstrated trajectories is identified as a central problem in LfD in [6]. In an attempt to improve extrapolation abilities of TP, [5] propose a modified Parametric Gaussian Mixture Model (mPGMM) that uses an altered Expectation Maximisation (EM) procedure to improve the extrapolation abilities of the learnt model. The authors show the proposed method improves the extrapolation abilities of TP models while reducing the computation time required compared to TP-GMR. While the proposed method helps in extrapolation of the demonstrated task, this method can fail to preserve local structure, resulting in failure to execute the desired task effectively.

Modification of the regression procedure is considered in [7], where the authors propose *confidence-weighted* task parameterised movement learning, but not in the context of *improving* extrapolation. The authors show how different trajectories can be generated from the same set of task parameters through weighting of the learned local models. The authors propose this is a useful mechanism for providing human prior knowledge about the importance of different task parameters to the model. While this is a possible use for the method, it seems non-trivial for a person to determine the relative importance of a task frame and manually assign weightings, particularly if that person is a novice user of the technology.

In [8], the authors discuss identifying *important* task parameters for the purpose of determining whether a defined task parameter is required to learn a task or not. Here, the magnitude of the covariance matrix, evaluated using a matrix determinant, is considered an indicator for this. Under Task-Parameterised Gaussian Mixture Model-based methods, variability in the recorded data is encoded in the local models' covariance matrices. Tighter groupings of data in the local models will result in "smaller" covariance matrices, hence the matrix determinant.

To address these issues, this paper builds on the confidence weighting scheme presented in [7] and the frame importance sampling in [8], and presents an *autonomous* method for weighting task frames during regression to improve model extrapolation, while retaining local structure observed in demonstration data.

## III. BACKGROUND

Central to understanding how the proposed method improves extrapolation is understanding how local structure is modelled and used in TP learning methods. To this end, a brief review of the related methods is presented here.

### A. Task Parameterised Learning

The learning process begins with the user providing a demonstration set consisting of  $M$  demonstrations, each containing  $T_m$  data points, collected in a global frame of reference. This

data is formed into a data set of  $N$  state measurements,  $\xi_n$ , with  $N = \sum_m^M T_m$ .

In addition to the raw data, Task Parameterised Learning builds local representations of the demonstrated task through a set of  $P$  task parameters. In their most general form, these task parameters are represented by sets of affine transformations, but in the context of this work they can be simply considered as coordinate frames of reference,

$$p_{n,j} = \{\mathbf{b}_n^{(j)}, \mathbf{A}_n^{(j)}\}, \quad (1)$$

with  $\mathbf{b}$  representing the location of the frame and  $\mathbf{A}$  representing the orientation of the coordinate frame.

The collected demonstrations are then projected into the local coordinate frames,

$$\mathbf{X}_n^{(j)} = (\mathbf{A}_n^{(j)})^{-1}(\xi_n - \mathbf{b}_n^{(j)}), \quad \mathbf{X}_n^{(j)} \in \mathbb{R}^{M \times N}, \quad (2)$$

where  $\mathbf{X}_n^{(j)}$  is the trajectory sample at time step  $n$  in frame  $j$ . Mixture models are then fitted to these local trajectory representations to build models of the local structure present in the data.

### B. Task-Parameterised Gaussian Mixture Models

Under Task-Parameterised Gaussian Mixture Model, a  $K$ -component mixture model is fit to the data in *each* frame of reference. Each GMM consists of mixing coefficients,  $\pi$ , means,  $\mu$ , and covariances,  $\Sigma$ . Together, these form a set of local mixture models, representing the task from multiple points of view,  $\{\pi_i, \{\mu_i^{(j)}, \Sigma_i^{(j)}\}_{j=1}^P\}_{i=1}^K$ .

To use the local models for trajectory generation they must first be projected back into the global frame of reference and then combined into one global model. This is achieved through a linear transformation of the local models with their respective task parameters, followed by a product of Gaussians,

$$\mathcal{N}(\hat{\xi}_{n,i}, \hat{\Sigma}_{n,i}) \propto \prod_{j=1}^P \mathcal{N}(\mathbf{A}_n^{(j)} \mu_i^{(j)} + \mathbf{b}_n^{(j)}, \mathbf{A}_n^{(j)} \Sigma_i^{(j)} \mathbf{A}_n^{(j)}) \quad (3)$$

With (3), if the same task parameters are used in the product as were used to learn the local models, the resulting mixture model will produce a trajectory that attempts to replicate a demonstration.

Generalisation with Task-Parameterised Gaussian Mixture Model also emerges from (3). That is, given new values for  $\{\mathbf{A}_j, \mathbf{b}_j\}$ , the local models can be used to generate global models for *different* task parameters. Effectively, the local models are placed in a new pose in task space when (3) employs new task parameters. Considering task parameters as representing end-points of a trajectory, or locations of objects, it is this linear transformation followed by a product of Gaussians that allows the local models to extrapolate to new situations.

Finally, Gaussian Mixture Regression (GMR) can be used to generate a smooth path from the global  $K$ -component model. Assuming the model is time-driven, the GMM will encode the joint probability of states and time,  $p(t, \xi_t)$ , from which states can be sampled through GMR that computes the conditional distribution  $p(\hat{\xi}_t|t)$ . See [4] for more details.

### C. Task Parameter Weightings

A key step in the proposed method is modifying the contribution of local models to the combined global model. A suitable method for this is proposed in [7], with the previously mentioned confidence weighted autonomy scheme. This involves scaling the covariance matrices of a local mixture model using a weighting parameter  $\alpha$ ,

$$\mathcal{N}(\hat{\xi}_{n,i}, \hat{\Sigma}_{n,i}) \propto \prod_{j=1}^P \mathcal{N}(\mathbf{A}_n^{(j)} \mu_i^{(j)} + \mathbf{b}_n^{(j)}, \mathbf{A}_n^{(j)} \Sigma_i^{(j)} / \alpha_n^{(j)} \mathbf{A}_n^{(j)}) \quad (4)$$

where  $\alpha_{n,j}$  is the weight value frame  $j$  at time step  $n$ , with the properties  $\alpha_{n,j} \in (0, 1)$ . As discussed in §II, it is proposed that these weight values can be used as a method for incorporating human prior knowledge to the model; however this may be challenging for novice users. Instead, a new weighting scheme based on the relevance of a local model is proposed.

### D. Frame Importance

As discussed in §II, the authors in [8] discuss identifying *important* task parameters. They define the importance of frame  $j$  at step  $n$ ,  $F_{n,j}$ , as the ratio of the precision matrix determinant for a given frame with respect to the other frames,

$$F_{n,j} = \frac{|\Sigma_{n,j}^{-1}|}{\sum_{j=1}^P |\Sigma_{n,j}^{-1}|}, \quad (5)$$

$$F_{n,j} \in (0, 1), \quad \sum_{j=1}^P F_{n,j} = 1 \quad \forall n.$$

This frame importance measure will form a first step in defining the frame relevance weightings of the proposed method.

## IV. METHOD

This section details the method used to determine optimal task parameter weightings for trajectory generation and improving the extrapolation ability of learned models. Optimisation of task parameter weightings is then shown through a task-independent, variance weighted cost function.

### A. Frame Relevance

By incorporating a weighting that captures frame importance at each sample along a trajectory, the goal is to allow the global model to generate trajectory points that only consider contributions from local models *when they are relevant* to the task, with the objective of improving task performance.

While the frame importance measure in [8] offers a possible solution to frame weighted trajectory generation, there are further steps that can be taken to ensure an optimal frame relevance weighting is selected.

First, the covariance matrices used in (5) are sampled from the learned model at the required time step through a GMR process. This has potential to introduce unwanted bias to the weightings, as a result of the choice of model parameters

such as number of Gaussian components. For the purpose of determining frame relevance as indicated by demonstrations, an alternative source of information is then to directly fit a single Gaussian at each time step to the data points in each local frame of reference

$$\{\xi_{m,n}^{(j)}\}_{m=1}^M \sim \mathcal{N}(\tilde{\mu}_n^{(j)}, \tilde{\Sigma}_n^{(j)}). \quad (6)$$

While this presents an additional computational cost, it is only required when the demonstration set is updated.

Next, instead of taking the inverse of the covariance matrices it is possible to parameterise this power,  $\gamma$ , such that it becomes possible to optimise the frame weightings for a particular task

$$\alpha_{n,j} = \frac{|(\tilde{\Sigma}_n^{(j)})^\gamma|}{\sum_{j=1}^P |(\tilde{\Sigma}_n^{(j)})^\gamma|}. \quad (7)$$

By selecting this parameterisation, the degree to which local correlations take precedence over global correlations is controllable. As  $\gamma$  scales, the covariance matrices will adjust as well, such that as  $\gamma$  is increased trajectory generation will tend to favour the local model structure of the “dominant” task parameter at a given time step. For example, in a grasping task as the robot approaches the object to grab it, the model will prioritise the model in the object’s local frame of reference. This is an important modification, as in a standard Task-Parameterised Gaussian Mixture Model the contribution of the start position frame in (3) will offset the trajectory slightly.

A final step taken to ensure smooth transition between local models during trajectory generation is applying a smoothing procedure to the resulting frame weightings using a moving average window.

### B. Optimising Frame Relevance Weights

Selection of  $\gamma$  can be optimised by treating the demonstration data set as a validation data set. Generating a set of trajectories using the task parameter sets from each demonstration will provide a set of trajectories that can be used to evaluate the learner’s performance on the demonstrated task.

A *weighted quadratic cost function* is defined, where the weights used directly model the variability in the data. This is done to prioritise parameter optimisation for regions of the trajectory where accuracy is required, as indicated by demonstration data variance. This is achieved by setting the weights to a diagonal matrix with entries equal to the norm of the generated data point’s covariance  $\Sigma$ , normalised such that it is in the range  $(0, 1)$ ,

$$\ell = \sum_{m=1}^M \sum_{n=1}^N (\xi_{d,m,n} - \xi_{g,m,n})^\top \mathbf{W}_{m,n} (\xi_{d,m,n} - \xi_{g,m,n}), \quad (8)$$

$$\mathbf{W}_{m,n} = \begin{bmatrix} \sigma_{m,n} & & \\ & \ddots & \\ & & \sigma_{m,n} \end{bmatrix}, \quad (9)$$

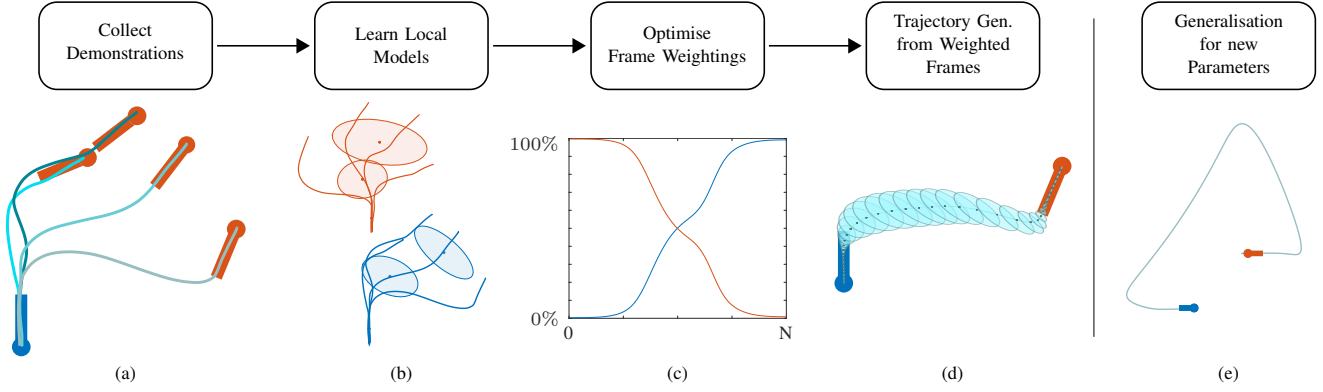


Fig. 1: Data processing and model learning pipeline of the  $\alpha$ TP-GMR.

$$\sigma_{m,n} = \frac{\|\Sigma_{m,n}\|}{\sum_{n=1}^N \|\Sigma_{m,n}\|}. \quad (10)$$

By defining the cost in this manner, the robot is able to prioritise its optimisation of  $\gamma$ , with higher costs being accumulated in regions of low variability (*i.e.*, high accuracy is required), and lower costs in regions of high variability.

The optimal value for  $\gamma$  can then be found through a one-dimensional parameter search method. This is achieved with a bounded golden section search method, as provided in *Matlab*.

The proposed method is summarised in Figure 1. This approach, with variance-adjusted frame weighting for trajectory optimisation, forms the Relevance-Weighted Task-Parameterised Gaussian Mixture Regression ( $\alpha$ TP-GMR) method.

## V. EXPERIMENTS

The proposed approach is first evaluated on a test data set for a reaching task, followed by performance evaluation on a real-world manipulation task with a more complex state representation. The objective in these experiments is to explore the proposed method’s ability to extrapolate tasks to unseen conditions, and show how this ability is of practical use in a real-world scenario.

### A. Reaching Task Performance

This first experiment investigates the performance of  $\alpha$ TP-GMR on a reaching task data set, compared to two contemporary methods, TP-GMR and mPGMM.

The reaching task data set used in this experiment<sup>1</sup>, consists of four demonstrations showing a point-to-point reaching task that approximates removing an end-effector from one pocket and inserting it into another (shown in Figure 1(a)). There are two sets of task parameters. The first parameter for each demonstration forms a coordinate frame centred on its start location, with the orientation aligned with the direction of travel. The second forms a coordinate frame centred on the goal location with a fixed orientation (red and blue markers in Figure 1 respectively).

<sup>1</sup>Reaching data set available from [http://www.idiap.ch/software/pbdlib/\[4\]](http://www.idiap.ch/software/pbdlib/[4])

1) *Setup*: In the data set, the state consists of a time index, and the location of the trajectory point,  $\xi_n = (t_n, x_n, y_n)^\top$ , where for the state at each sample  $n$ ,  $t$  is the time step. The task frames,  $\{\mathbf{b}, \mathbf{A}\}$ , are then defined as follows,

$$\mathbf{b}_{m,p} = (0, x_{m,p}, y_{m,p})^\top, \quad \mathbf{A}_{m,p} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{m,p} \end{pmatrix}, \quad (11)$$

where  $(x_m, y_m)$  is the position of the  $p^{th}$  frame for the  $m^{th}$  demonstration, and  $\mathbf{R}_{m,p} \in \mathbb{R}^{2 \times 2}$  is a planar rotation matrix representing the orientation of the frame. The task frames are static over time steps, but vary per demonstration  $m$ .

In addition to task parameters,  $K = 3$  components are used for each of the models. For  $\alpha$ TP-GMR, confidence weightings are estimated for the frames following the procedure described in §IV.

Evaluation of a model’s ability to learn the task performance is achieved through an exhaustive *leave-one-out* cross-validation procedure. For each model option, TP-GMR, mPGMM, and  $\alpha$ TP-GMR, a model is learned using  $M - 1$  of the available demonstrations. The reproduction score for the selected model is then taken as the Root Mean Square Error (RMSE) between the set-aside trajectory and a trajectory generated from the learnt model with the remaining set-aside trajectory’s task parameters. This procedure is repeated, cycling which demonstration is left out and resetting the model on each attempt, until each demonstration has been used as the cross-validation test trajectory.

2) *Results & Discussion*: Table I shows the results from this initial test with the point-to-point reaching data set.

It can be seen that of the three methods tested,  $\alpha$ TP-GMR incurs the lowest error for this data set. This indicates that  $\alpha$ TP-GMR was the most accurate in generating trajectories for unseen conditions, albeit over a small range.

TABLE I: Exhaustive leave-1-out cross-validation results for §V-A.

	TP-GMR	$\alpha$ TP-GMR	mPGMM
RMSE	0.279	0.197	0.270
Std.	$\pm 0.146$	$\pm 0.112$	$\pm 0.140$

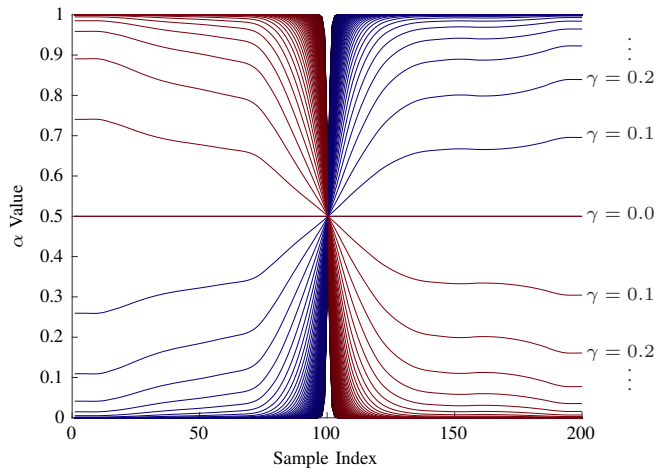


Fig. 2: Effect of  $\gamma$  on  $\alpha$  for the two-frame model used in §V-A. Lines indicate frame relevance over the course of a trajectory from the demonstration test data set, with red representing Frame 1, and blue Frame 2.

Looking more closely at the learning process for  $\alpha$ TP-GMR, Figure 2 shows plots of frame weightings generated for the reaching task model, with increasing values of  $\gamma$ . Here, the red lines indicate the start frame, and the blue lines indicate the goal frame. As a frame weighting value increases, it decreases the corresponding covariance as defined in (4), as expected. Initially, the first frame takes priority followed by a transition to the second frame as the trajectory approaches the goal. The key observation here is that, for increasing  $\gamma$ , the frame weighting will increasingly favour one local model over the other. At  $\gamma = 0$ , each frame is given equal weighting  $\alpha_j = 0.5$ ; however as  $\gamma$  increases, the transition from one frame to another becomes increasingly steep. It is the controllability of this transition that allows the  $\alpha$ TP-GMR method to optimise trajectory generation effectively.

TABLE II: Summary statistics for grid-search test case where the starting frames location is varied, but its rotation is similar to rotations observed in the demonstration set.

	Constraint Errors		Task Errors		Path Lengths	
	Mean	Std.	Mean	Std.	Mean	Std.
TP-GMR	19.34	$\pm 2.59$	1.10	$\pm 0.74$	9.53	$\pm 3.68$
$\alpha$ TP-GMR	1.00	$\pm 0.00$	0.04	$\pm 0.00$	9.73	$\pm 3.40$
mPGMM	18.00	$\pm 2.76$	1.03	$\pm 1.30$	10.66	$\pm 4.84$

### B. Extrapolation Performance

The second experiment investigates the method’s ability to extrapolate task performance to unseen conditions. This is shown through a grid search approach that expands far around the original demonstration area, where trajectories are generated with the learnt model for a series of starting positions.

1) *Setup*: Here, a  $10\text{m} \times 10\text{m}$  grid of task parameters is constructed, centred on  $(0, 0)$ . Each parameter in the grid is given an orientation similar to those encountered in the demonstration set. Each of these start position parameters is

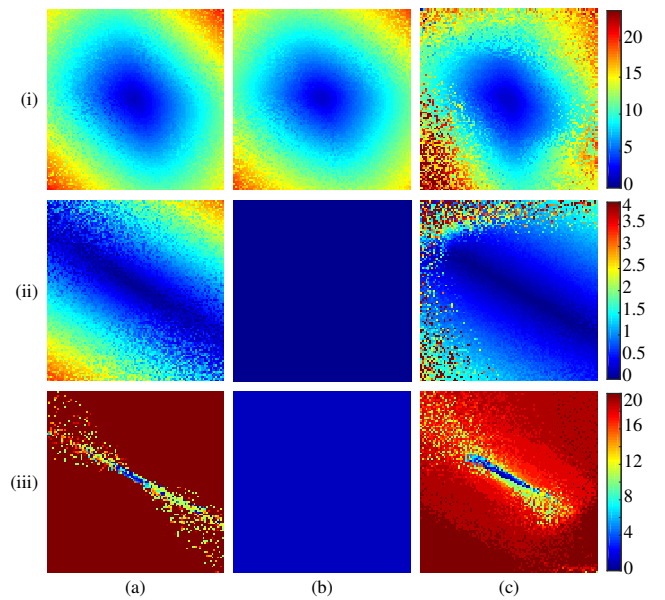


Fig. 3: Generalisation test results for three learning methods. Column (a) TP-GMR, (b)  $\alpha$ TP-GMR, (c) mPGMM. Row (i) total path length, (ii) Euclidean distance error at start and end of trajectory, (iii) Constraint satisfaction at start and end of trajectory.

then paired with a goal parameter that is the same as the one used in the demonstrations. A large degree of extrapolation is considered here. In the original data set, the goal location is set at  $(-0.8, -0.8)$  and each of the start task parameters are located  $\sim 1.5\text{m}$  away from the goal.

For each parameter set in the grid, a trajectory is generated and evaluated on three criteria. The criteria are (i) trajectory length, (ii) trajectory end-points error, (iii) constraint satisfaction error. Trajectory length is taken as an indicator of the quality of the demonstration, as longer trajectories can be an indicator of incoherent paths and shorter trajectories can be an indicator of incomplete paths. Trajectory end-point error specifically evaluates the model’s ability to generate a trajectory that starts and ends where it is meant to. As found in [4], task modelling methods that do not exploit local structure can rapidly see this type of error increase, potentially resulting in erratic movement at the start of a trajectory and incomplete actions due to movements ending early.

The constraint satisfaction error is designed to capture the model’s ability to generate paths that exit and enter the start and end frames in the correct orientation. From the start frame of reference, trajectories should move in the direction that the frame is pointing until the path is clear of the frame marker. For the goal frame of reference, trajectories should enter directly down from the top of the frame marker. Pragmatically, this is evaluated by counting how many data points are within bounding boxes placed at each end of the generated trajectory. These bounding boxes are chosen such that the first and last 10 data points of each demonstration trajectory are contained within them. The error count is taken as absolute value, so that the learning method will



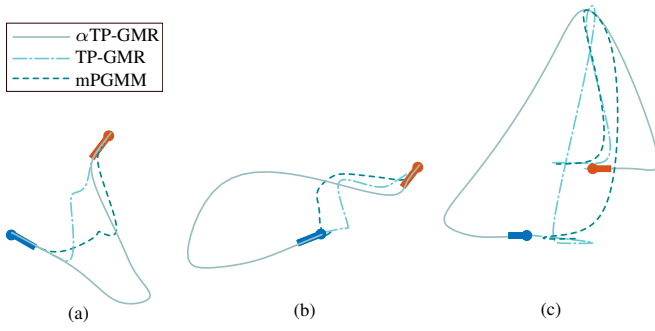


Fig. 4: Extrapolation tests comparing the proposed method against alternative models under three conditions. (a) Goal frame rotated 120 degrees, (b) Goal frame rotated 240 degrees, (c) Goal and origin frame rotated 90 and 270 degrees respectively. The generated trajectories must start and end at the red and blue frames respectively, with the paths exiting and entering along the direction indicated by the frame indicator. Note that only the proposed method succeeds in all cases.

be penalised for too many as well as too few data points being located in these bounding boxes. If 10 data points are counted in each bounding box, it is assumed that the trajectory satisfied the task constraint.

## 2) Results & Discussion:

The results in Table II highlight the significant difference in performance between the three methods. Here, it can be seen that  $\alpha$ TP-GMR achieves much lower task and constraint error values. This indicates that under  $\alpha$ TP-GMR, the model is able to generate trajectories that accurately produce the requested path, and importantly this path follows local structure constraints, as provided in the original demonstrates.

Plots of the results are provided in Figure 3, where each method occupies one column, and each criteria is plotted along one row. The distance criteria is plotted along row (i). It can be seen that for (a) TP-GMR and (b)  $\alpha$ TP-GMR that the generated path lengths are largely similar; however for (b) mPGMM, the trajectories begin to become erratic in the extremes of the grid.

Looking at row (ii), end-point error, it can be seen that the performance of (a) TP-GMR degrades in a regular pattern as trajectories move further away from the original demonstration set. It can be seen that mPGMM provides improved performance over TP-GMR; however in the upper and lower left corners the trajectory generation becomes erratic in a manner similar to the first row. In addition to erratic end-point error, it can be seen that the error in (c) does increase with a regular pattern like (a), albeit to a lesser degree. Plot (b) presents the first unusual result, where it can be seen that the error for  $\alpha$ TP-GMR is very low and constant across the grid. This result is made possible by the clean data in the data set producing frame weightings that accurately prioritise one frame over the other.

Looking at the final row in Figure 3, task constraint errors, reveals another unusual result for  $\alpha$ TP-GMR. In (a), TP-GMR can be seen to have a very small low-error region, which directly lines-up with the original demonstration region. From this, it can be concluded that TP-GMR is only effective in the neighbourhood of the original demonstrations, given the patterns seen in end-point error and



Fig. 5: Horticultural sorting task used in §V-C. Here, unhealthy plants must be removed from a plant tray. Key to this is the ability to generalise from user provided demonstrated trajectories (cyan), to determine the appropriate trajectory to pick plants from previously unseen locations (red).

constraint error. In (c), mPGMM does not fare much better than TP-GMR and similar conclusions can be drawn. Finally in (b), it can be seen that again there is a low, constant level of error across the grid.

The combination of low end-point error, low task-constraint error, and a smoothly increasing path length is a powerful combination. While these strong results are largely due to the clean nature of the data set, they are indicative of the ability for  $\alpha$ TP-GMR to greatly enhance the extrapolative abilities of LfD systems.

Figure 4 presents some samples from each of the models in a variety of generalisation challenges. It can be seen that in each case,  $\alpha$ TP-GMR is able to generate a smooth trajectory which satisfies the task constraints.

These results raise the question of whether an extrapolated trajectory is correct and should be used. If the model can produce an accurate trajectory to a previously unseen scenario, there is uncertainty over whether this is a safe or correct action to take. Some steps to automate detection of uncertain states can be found in [9]; however whether or not to trust the system largely remains at the discretion of the user. Ultimately, if a trajectory is suitably extrapolated, and the person agrees, this presents a large time saving for them.

## C. Real World Manipulation Task

Having seen the benefits of the proposed method on the reaching task data set, a real-world task is considered that presents further challenges.

In this experiment, the data used was collected from a real-world robot system, with demonstrations provided by *novice users* (i.e., people who do not have prior knowledge

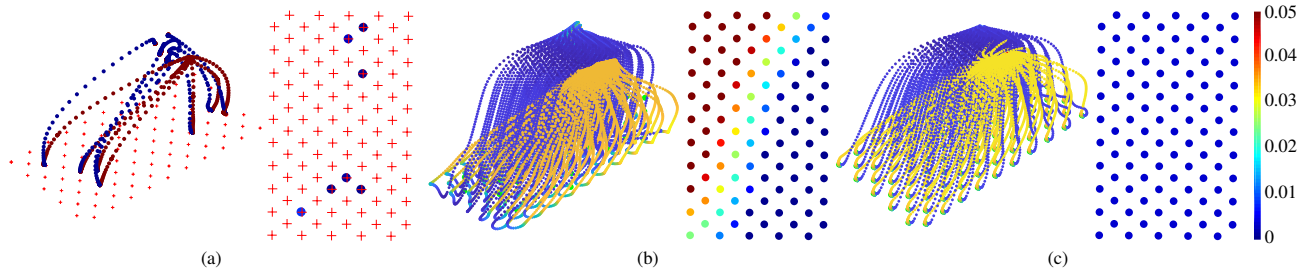


Fig. 6: Demonstration data (a) provided by novice users on the real world horticultural task shown in Figure 5. (b) Displays the trajectories generated using TP-GMR. (c) Displays the trajectories generated using  $\alpha$ TP-GMR. The colour of the dots in the grid plots indicates the positioning error of the trajectory when executing a grab action for that dot. The color of the trajectories indicates the hand control signal state, with the displayed plots showing the gripper beginning open and closing once it is at the grasp target.

of robotics, machine learning, LfD, *etc.*)<sup>23</sup>. The robot used is a Rethink Robotics Sawyer, with an Active8 AR10 hand, and the task under consideration is a horticultural sorting task as found in mass production sites of ornamental plants, where rejected products must be removed from a tray and discarded, see Figure 5.

LfD is useful for this task, as there can be a great deal of variety in the production process on grower sites. There can be hundreds of varieties of plants grown at one site over the course of a year, with different plants requiring different manipulation strategies. Additionally, the plants and flowers are grown in a variety of pots and tray ranging in capacity from 25-100 plants. In this scenario, a learning method that is able to *accurately* generalise from a few demonstrations to many locations would be a great help in providing flexible automation.

1) *Setup*: In this experiment, TP-GMR and  $\alpha$ TP-GMR are used to learn models of the task. Specifically, the task involves learning to pick up, remove, and place a plant from a tray of 100 to a disposal bucket. The objective provided to the participants was to teach the robot to perform the disposal task for any plant position in the tray. This task was demonstrated by 36 participants, 3 times each, providing 108 teaching interactions to consider.

The state and task parameters are then defined as follows,

$$\xi_n = \begin{bmatrix} t_n \\ \mathbf{x}_n^p \\ \mathbf{x}_n^q \\ x_n^h \end{bmatrix} \in \mathbb{R}^{8 \times 1}, \quad (12)$$

$$\mathbf{b}_{m,p} = \begin{bmatrix} 0 \\ \mathbf{p}_{m,j} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{8 \times 1}, \mathbf{A}_{m,p} = \mathbf{I}^{8 \times 8} \quad (13)$$

where  $\mathbf{x}_n^p$  and  $\mathbf{p}_{m,j}$  are the positions of data point  $n$  and frame  $j$  for demonstration  $m$ , respectively, and  $\mathbf{x}_n^q$  is the orientation of data point  $n$  demonstration  $m$  using an axis-angle representation.  $x_n^h$  is a scalar control signal used to

open and close the robot hand, and  $\mathbf{I}^{8 \times 8}$  is the identity matrix.

The performance of the robot learner under the two learning methods is then evaluated by generating a test set of trajectories for each of the 100 plant positions in the tray. Each trajectory is evaluated by comparing the end-effector position at critical points during the task execution.

These critical points are (i) the start location, (ii) the grab location, and (iii) the place location. Position (i) ensures that the model is generating a trajectory that starts where it is meant to, thus avoiding sudden jerks in movement at the start of the task. Position (ii) is evaluated by identifying the location of the robot hand at the point it closes its gripper, and comparing it to the mean location of the hand during grasping in the demonstration set. Position (iii) then ensures that the robot is correctly depositing the picked plant, and is identified as the point at which the robot opens its hand.

In evaluating the learning methods in this way, it is assumed that the demonstration data provides correct information on how to pick the plants from the tray, and that if the local structure of the generated trajectory (*i.e.*, the grabbing location relative to the plant) does not closely match the demonstration data, then the robot will fail to grasp the plant.

2) *Results & Discussion*: Analysing the collected data revealed the distribution of data residuals was found to be non-normal by an Anderson-Darling test on the data. Considering this non-normality, the data was tested using a paired non-parametric Wilcoxon Signed-Rank test. This indicated that the median error  $\alpha$ TP-GMR ( $\tilde{\epsilon} = 5.0541$ ) was statistically significantly lower than under conventional TP-GMR ( $\tilde{\epsilon} = 3.4763$ ),  $Z = 8.6621$ ,  $p < 10^{-17}$ .

This result confirms the findings observed in the previous two experiments. Note that in all experiments, while the learning methods were assessed using task-specific criteria, the learning process used the task-independent cost function (8) described in §IV-B.

Looking more closely at the results and plotting a selection of the data generated by TP-GMR and  $\alpha$ TP-GMR reveals some useful details. A demonstration set is shown in Figure 6(a) along with its corresponding grasp-error plot. When using TP-GMR to learn from this data set, as shown in Figure 6(b), this data set can be seen to be suboptimal. There

<sup>2</sup>This experiment was conducted with ethical approval granted by KCL REC Committee under LRS-17/18-5549.

<sup>3</sup>A DOI to this data set will be made available in the final paper.



are two sub-groups of demonstrations in the top and bottom portion of the tray, which results in redundant demonstrations in the demonstrated regions, and undemonstrated states elsewhere. By switching to a  $\alpha$ TP-GMR learning mode, with no adjustment to the demonstration set, Figure 6(c) shows a large improvement in the performance of the learner. In particular, it can be seen that in Figure 6, the trajectories near the grasp targets closely match the trajectories in the demonstration set, indicating that the local structure has been learned and is being used.

This is an important result in LfD. Given that people often struggle to provide adequate demonstration sets [1], [3], a learning method that can effectively extrapolate from regions that have been shown could reduce the challenge of providing good demonstration sets for LfD.

Note that unlike in the first set of reaching task experiments, the error achieved does not reduce to a constant level. This is due to the noisy nature of the data making learning more challenging, and resulting in less information being available to the robot learner to gauge which frames are important at each step.

## VI. CONCLUSION

Extrapolation in LfD presents many challenges and opportunities. As discussed in related work, and shown through experiments in §V-A, prior approaches to improving extrapolation in Task Parameterised learning have limited generalisation abilities beyond the original demonstrations. This paper presents a new approach, Relevance-Weighted Task-Parameterised Gaussian Mixture Regression, for addressing this problem. Under this method, task parameters are modulated based on their estimated importance during each time step in a trajectory.

As demonstrated in a series of experiments with both simulated data for a reaching task, and real world data collected from novice users, this approach significantly improves the extrapolation abilities of TP-GMR. These improvements will serve to benefit novice users of LfD systems, by enhancing the ability of robot learners to extrapolate from limited data and places less reliance on the user providing a high-quality demonstration data sets.

Limitations of the proposed approach include the issue of time distortion. Generated trajectories have the same number of data points as original demonstrations, so additional processing may be required to ensure robot limits are not exceeded. There is also the more fundamental question of how to determine whether an extrapolated trajectory is correct and should be trusted. Whilst this is a common consideration when extrapolating using any learning method, the high degree of extrapolation possible with  $\alpha$ TP-GMR might mislead a novice user to be confident in a generated trajectory that is unsuitable.

Future work could consider the effect of  $\alpha$ TP-GMR on users presented with extrapolated trajectories. Further, the discussed models represent a small subset of models available in TP. It would be of interest to explore how relevance

based frame weighting can be applied to state-based systems, and tasks with force interactions.

## ACKNOWLEDGMENTS

The authors wish to thank the many people who took time to participate in this experiment.

## REFERENCES

- [1] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009.
- [2] S. Calinon and A. Billard, "What is the Teacher's Role in Robot Programming by Demonstration? - Toward Benchmarks for Improved Learning," *Interaction Studies. Special Issue on Psychological Benchmarks in Human-Robot Interaction*, vol. 8, no. 3, 2007.
- [3] A. Sena, Y. Zhao, and M. J. Howard, "Teaching Human Teachers to Teach Robot Learners," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5 2018, pp. 1–7.
- [4] S. Calinon, "A tutorial on task-parameterized movement learning and retrieval," *Intelligent Service Robotics*, vol. 9, no. 1, pp. 1–29, 9 2015.
- [5] S. Calinon, T. Alizadeh, and D. G. Caldwell, "On improving the extrapolation capability of task-parameterized movement models," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 11 2013, pp. 610–616.
- [6] T. Osa, J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel, and J. Peters, "An Algorithmic Perspective on Imitation Learning," *Foundations and Trends in Robotics*, vol. 7, no. 1-2, pp. 1–179, 2018.
- [7] Y. Huang, J. Silverio, L. Roza, and D. G. Caldwell, "Generalized Task-Parameterized Skill Learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5 2018, pp. 1–5.
- [8] T. Alizadeh, S. Calinon, and D. G. Caldwell, "Learning from demonstrations with partially observable task parameters," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5 2014, pp. 3309–3314.
- [9] G. Maeda, M. Ewerton, T. Osa, B. Busch, and J. Peters, "Active Incremental Learning of Robot Movement Primitives," in *Conference on Robot Learning*. PMLR, 2017.